Chua, Patric, Har, Zi Mei, Austin, Christopher M., Yule, Catherine M., Dykes, Gary A. and Lee, Sui Mae 2015, Genome sequencing and annotation of *Cellulomonas* sp. HZM, *Genomics data*, vol. 5, pp. 40-41.

DOI: 10.1016/j.gdata.2015.05.009

This is the published version.

©2015, The Authors

Reproduced by Deakin University under the terms of the Creative Commons Attribution Non-Commercial No-Derivatives Licence

Available from Deakin Research Online:

http://hdl.handle.net/10536/DRO/DU:30091309
Data in Brief

Genome sequencing and annotation of *Cellulomonas* sp. HZM

Patric Chua *, Zi Mei Har, Christopher M. Austin, Catherine M. Yule, Gary A. Dykes, Sui Mae Lee

School of Science, Monash University, Selangor, Malaysia

**Abstract**

We report the draft genome sequence of *Cellulomonas* sp. HZM, isolated from a tropical peat swamp forest. The draft genome size is 3,559,280 bp with a G + C content of 73% and contains 3 rRNA sequences (single copies of 5S, 16S and 23S rRNA).

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

**Keywords:**
- Tropical peat swamp
- *Cellulomonas*
- Cellulolytic
- Whole genome sequencing

**Specifications**

<table>
<thead>
<tr>
<th>Organism/cell line/tissue</th>
<th><em>Cellulomonas</em> sp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strain(s)</td>
<td>HZM</td>
</tr>
<tr>
<td>Sequencer or array type</td>
<td>Sequencer; Illumina MiSeq</td>
</tr>
<tr>
<td>Data format</td>
<td>Processed</td>
</tr>
<tr>
<td>Experimental factors</td>
<td>Microbial strains</td>
</tr>
<tr>
<td>Experimental features</td>
<td>Draft genome sequence of <em>Cellulomonas</em> sp.</td>
</tr>
<tr>
<td>Consent</td>
<td>N/A</td>
</tr>
<tr>
<td>Sample source location</td>
<td>Tropical peat swamp in Pekan, Pahang, Malaysia</td>
</tr>
</tbody>
</table>

**1. Direct link to deposited data**


**2. Experimental design, materials and methods**

*Cellulomonas* sp. HZM is a Gram positive, rod-shaped bacterium isolated from soil surface plant detritus in Pekan tropical peat swamp in Pahang, Malaysia. The isolate was acquired by culture-plating on Sizova’s cellulose minimal salt media [1] preceded by a series of enrichment steps. Genomic DNA was extracted from 24 hour old cultures using the GF−1 nucleic acid extraction kit (Vivantis, Malaysia) as per manufacturer’s instructions. The genome was sequenced using an Illumina MiSeq sequencer (150-bp paired-end reads) and the raw reads were trimmed and assembled de novo using CLC Genomics Workbench 6 (CLC Bio, Denmark). A total of 78 contigs with an accumulated length of 3,559,280 bp (74-fold coverage, N50 = 99,115 bp) and a G + C content of 73% were obtained. The genome contains 52 tRNA genes and 3 rRNA genes (5S–23S–16S) as predicted using tRNAscan 1.2 [2] and RNAmmer 1.2 [3], respectively.

A total of 3180 coding sequences in 352 subsystems were functionally annotated by Rapid Annotation using the Subsystems Technology (RAST) server [Fig. 1.](#)

1. Functional comparison of genome sequences in the RAST server revealed the closest neighbors of *Cellulomonas* sp. HZM as *Sanguibacter keddii* DSM 10,542 (score 555) followed by *Cellulomonas flavigena* DZM 20,109 (score 335), *Beutenbergia cavernae* DSM 12,333 (score 293), *Actinomyces odontolyticus* ATCC 17,982 (score 189) and *Xylanimonas cellulosilytica* DSM 15,894 (score 186). Analysis using the complete 16S rRNA sequence (performed using EzTaxon at [http://www.ezbiocloud.net/eztaxon](http://www.ezbiocloud.net/eztaxon)) is more specific by comparison, with the list of candidates being predominantly *Cellulomonas*, whereby the closest match is *Cellulomonas uda* (pairwise similarity of 98.20%)

---

* Corresponding author.
E-mail address: tcchu4@student.monash.edu (P. Chua).

http://dx.doi.org/10.1016/j.gdata.2015.05.009
2213-5960/© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
followed by *Cellulomonas chitinilytica* (98.15%), *Cellulomonas gelida* (98.06%) and *Cellulomonas iranensis* (97.89%).

3. Nucleotide sequence accession number

The whole genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. JEOE00000000.

Conflict of interest

The authors declare that there is no conflict of interests with respect to the work published in this paper.

Acknowledgments

Funding for this study was provided by the School of Science, Monash University Malaysia. We are grateful to the Pahang State Forestry Department for their assistance in procuring the samples.

**References**


